Тесты на случайность. Тесты, основанные на общем количестве серий.

На семинаре мы продолжили построение статистического критерия, определяющего случайность выборки по общему числу серий R в последовательности, составленной из элементов двух типов.

Моменты R

По определению,

$$\mathbf{E}R^k = \sum_r r^k f_R(r) = \left\{ \sum_{r \text{ чёт}} r^k f_1(r) + \sum_{r \text{ нечёт}} r^k f_2(r) \right\}$$

Вычисление моментов по этой формуле — трудоемкий процесс. Более простой способ для вычисления моментов — представить R в следующем виде

$$R = 1 + I_2 + I_3 + \ldots + I_n,$$

где

$$I_k = \begin{cases} 1, & \text{если } k\text{-ый элемент не равен } (k-1)\text{-му}, \\ 0, & \text{иначе.} \end{cases}$$

Тогда I_k — бернуллиевские случайные величины с параметром $p = \frac{n_1 n_2}{\binom{n}{2}}$, следовательно

$$\mathbf{E}I_k = \mathbf{E}I_k^2 = \frac{2n_1 n_2}{n(n-1)}.$$

Тогда

$$\mathbf{E}R = 1 + \frac{2n_1 n_2}{n_1 + n_2}, \quad \mathbf{D}R = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

Асимптотическое распределение R при верной H_0

Формула для $f_R(r)$ может быть использована для всех значений n_1, n_2. Вычисления трудоемки, если n_1, n_2 большие. Для больших выборок можно использовать приближение нулевого распределения R.

Для поиска асимптотики распределения предполагаем, что выборка имеет размер $n \to \infty$ и $n_1/n \to \lambda$, $n_2/n \to 1-\lambda$, где $\lambda \in [0,1]$ — фиксированное число. При больших размерах выборки

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{R}{n}\right) = 2\lambda(1-\lambda), \quad \lim_{n\to\infty} \mathbf{E}\left(\frac{R}{\sqrt{n}}\right) = 4\lambda^2(1-\lambda)^2.$$

Сформируем случайную величину

$$Z = \frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)}.$$

Подставив R в терминах Z в $f_R(r)$, получим $f_Z(z)$. Для неё

$$\lim_{n\to\infty} \ln f_Z(z) = -\ln\sqrt{2\pi} - \frac{1}{2}z^2,$$

т.е. предельная функция вероятности Z является плотностью стандартного нормального распределения.

Для двустороннего теста размера α нулевая гипотеза случайности H_0 будет отвергнута, когда

$$\left|\frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)}\right| \geq z_{\alpha/2},$$

где z_γ — $(1-\gamma)$-квантиль.